of the utilization rates (discharges age 65 and over). The negative correlation of hospital utilization with percent females age 15-44 is probably because an area with a high percentage of females age 15-44 will have a relatively young population, which has low utilization. This will counteract the higher number of obstetric discharges from this group, since obstetric discharges are a small percent of the total.

Factor Analysis

The factor analysis, which was intended to group the independent variables into a smaller set of uncorrelated "factors" to use in regression analysis, produced five major interpretable factors: medical resource factor, poverty factor, elderly factor, rural factor, and a white-collar factor. The prediction of hospital utilization using these factors was, however, rather poor with only 11 percent of the variation in total hospital utilization rates accounted for. A possible explanation for this is that the factor analysis grouped independent variables based on their correlations among themselves, not on their correlations with the dependent variables. In general, however, factor analysis should be considered as a technique for reducing a set of intercorrelated variables prior to regression analysis.

Regression Analysis

Table 2 exhibits the results of multiple regression analysis. The ten best variables were first chosen by stepwise regression and then entered into a multiple regression procedure that generated standardized weights for the independent variables. (In the SAS computer program that was used, the stepwise procedure does not produce standardized weights.) These weights indicate how much change in the dependent variable is produced by a change in one of the independent variables when the others are statistically held constant. Since these weights are standardized, we can rank the variables by their ability to predict utilization, after the variance shared with all of the other independent variables in the equation has been removed. The results from Table 2, therefore, indicate that the bed-to-population ratio is the best single predictor of utilization in this analysis, for total utilization and two of the age groups. The higher the beds per population in a county, the higher is the utilization of hospitals by residents of the county. Another very important variable is the physician-to-population ratio, which has a high, **negative** association with three of the four utilization rates, meaning that a low number of physicians per 1,000 population is related to a high rate of hospital utilization.

Other findings in Table 2 include the consistently positive effect on utilization of the percent of the population that are Medicare disabled enrollees. As was stated before, the percent Medicare disabled enrollees could be highly correlated with other variables that have a more direct impact on utilization. Average length of stay, meanwhile, has a consistently inverse relationship to the utilization rates. This negative relationship was also found in the correlation analysis.

The negative weight for the percent Medicare of total resident patients is rather curious since the simple correlation between this variable and utilization is positive. Thus, even though an area with high hospital utilization is likely to have a high percentage of Medicare patients, the contribution of this variable to the prediction of utilization in the multiple regression equation is negative. This paradox results from the fact that in multiple regression the sign of the weight applies to the relationship between that portion of hospital utilization unexplained by the other nine variables and that portion of the percent Medicare of total resident patients that is unrelated to these other nine variables. In this case, the Medicare percentage is positively related to some of the other variables in the equation and the positive explanatory effect of the Medicare percentage has apparently already been "used up" by these other variables. The weights of some other variables in Table 2 may have signs that run counter to "common sense." But again, keep in mind that multiple regression maximizes the power of a **combination** of variables to predict utilization, and this may produce results different from what a series of two-variable relationships would indicate.

As was noted in the methodology section, the R^2 statistic indicates the proportion of the total variance in the dependent variable that is accounted for by all of the independent variables. Thus we can account for around 40 percent of the variation in discharges age 0-13 per 1,000 population and 55 percent of the variance in the age 65 and over rates. For ages 14-64 and the total discharge rate, over 60 percent of the variance is accounted for by the regression equations, which is a reasonably good level of prediction.

The F value in Table 2 measures the significance of R^2. It is a test to see if the R^2 may be due simply to random variation. The probability statistic (p) represents the probability of getting the observed F value simply by chance. In all four multiple regression equations, the probability statistic is 0.0001 or lower which means that